An Ensemble of Machine Learning Models for Multilabel Classification of Cardiovascular Diseases by ECGs

Anastasia Bazhutina[1], Svyatoslav Khamzin[1], Alexander Sinitca[1], Mikhail Chmelevsky[1,2,3], Stepan Zubarev[1,3], Margarita Budanova[1,3], Werner Rainer[1]

[1]XSpline S.p.A, Bolzano, Italy
[2]Division of Cardiology, Fondazione Cardiocentro Ticino, Lugano, Switzerland
[3]Almazov National Medical Research Centre, Saint-Petersburg, Russia

Abstract

The study focuses on automating ECG analysis, a crucial tool for cardiac evaluation and treatment decisions. Two classifiers were developed to detect cardiac conduction disorders and myocardial infarctions from 12-lead ECGs. These classifiers were trained on three open datasets after ECG preprocessing to remove noise and select relevant features.

The first classifier (10 classes) identifies conditions of the ventricular conduction system: normal ECG, LBBB, ILBBB, premature ventricular complex, LAFB, LPFB, IRBBB, RBBB, nonspecific intraventricular conduction disturbance, and ventricular preexcitation. The second classifier (5 classes) distinguishes between non-infarcted ECGs and various types of myocardial infarction locations.

For each QRS complex, we calculated statistical, time-domain, frequency features, and scalograms. Then we trained classifiers combining XGBoost for statistical features, an autoencoder+neural network for deep-learning features, and ResNet for scalograms. An ensemble approach was employed to get the final class prediction. We show that the developed ensemble model achieve the mean F1-score of 0.70 and 0.76 for the first and second classifier respectively. The study's key strength lies in its diverse feature extraction methods, enhancing the predictive power of machine learning models.

1. Introduction

The recognition of abnormalities in the electrocardiograms (ECGs) by computers is widely used by cardiologists to categorize long-term ECG recordings. The combination of Machine Learning (ML) models and ECGs has opened up exciting possibilities for improving the accuracy and efficiency of these diagnostics.

Various feature extraction techniques have been employed in this context, including morphological features [1], frequency-domain features [2], complex heartbeat representations [3], wavelet-based features [4], and statistical features [5].

These extracted features can then be classified using a range of methodologies. The proposed techniques encompass simple classifiers like linear discriminants (LD) [6], nearest neighbor rules [7], and decision trees [8], as well as more advanced approaches such as neural networks [9], conditional random fields [10], and, more recently, deep learning techniques [11].

The main difficulty in developing such models is that multiple labels can be assigned to a single ECG recording. Thus, the requirements for the training dataset, which must contain a sufficient number of rare cardiac disease combinations, are increased.

In this study, we proposed an ensemble of ML models for the classification of ECG diseases. For this preliminary study, we selected only QRS features to classify heart ventricle diseases. We propose different ways to extract ECG features and then use an ensemble of ML models to finalize classification.

2. Methods

In this work, we built two ML classifiers to determine the diagnosis from the ECG - a **10-class** classifier and a **5-class** classifier. The pipeline of this study is presented in Figure 1.

For the **10-class** classifier, we selected the following labels: normal ECG, left bundle branch block (LBBB), incomplete left bundle branch block (ILBBB), premature ventricular complex (PVCs), left anterior fascicular block (LAFB), left posterior fascicular block (LPFB), incomplete right bundle branch block (IRBBB), right bundle branch block (RBBB), non-specific intraventricular conduction disturbance (NICD), and ventricular preexcitation (VPE) (see Table 1). For the **5-class** classifier, we selected

the following labels: non-infarcted ECGs, anteroseptal myocardial infarction (MI), lateral MI, inferior MI, and anterior MI (see Table 2).

2.1. Initial data

To train and validate our classifiers, we used three 12-lead datasets: the dataset of Chapman University [12], the PTB-XL dataset [13], and the dataset of Shandong Provincial Hospital (SPH) [14]. The datasets were collected from different clinics, each containing a different number of labels and records. Moreover, there is a significant imbalance in the data, which makes the classification task challenging. In addition, only 12-lead ECG recordings with a frequency of 100 out of 500 Hz were used. The ECG recordings were divided by a ratio of 0.8 to 0.2 for training and validation data, respectively.

2.2. ECG preprocessing

At the initial stage, the ECG data has been filtered and any unsuitable signals have been excluded. The signals have been further processed to remove baseline wanders, along with other types of noise. We used a moving average filter to remove baseline wander and to remove other types of noise, such as power line noise, muscle noise, and respiration noise bandpass filter was used, with a range of 0.1 50 Hz.

Then, we estimated the boundaries of the QRS complex in the patients' ECG signals using the Hamilton-Thompson algorithm [15]. To standardize the inputs for the neural network, we cut off the signal with the QRS complex within the boundaries.

2.3. Statistical features classification

For one of the ensemble ML models, we used statistical features and the XGBoost model. A similar approach was proposed in a recent article [5]. The mean and median features were utilized to determine the central tendency of the ECG signal. To capture the statistical dispersion of the ECG features such as standard deviation, range, and interquartile range were employed. The kurtosis and skewness parameters are employed to determine the degree of asymmetry and peakedness of the ECG signal distribution. So, we extracted statistical features from every lead signal and used a concatenated vector of features to classify ECG using the XGBoost classifier.

Additionally, statistical features were used as an additional step of ECG filtering. ECG signals with statistical features beyond 3 sigma were excluded from training and validation.

2.4. Wavelet features classification

To train the deep learning model, we used a comprehensive approach that involved transforming 12-lead ECG signals into complex 3D representations. This transformation was achieved by using the Continuous Wavelet Transformation (CWT) technique, which uses a Morlet mother wavelet as the basis. By applying the CWT, we generated scalograms [16], which are essentially scaled representations of the ECG signals. These scalograms took the form of 3D images, incorporating both time and frequency dimensions.

These 3D images were then used as the primary input data for a Residual Neural Network (ResNet) [17]. In the final stage of our model architecture, we implemented a fully connected layer with the Sigmoid as an activation function. Depending on the specific classification task, we used either a linear layer with 10 outputs for a 10-class classification problem or a linear layer with 5 outputs for a 5-class classification problem. The ResNet was used to solve the multi-label classification problem for ECGs.

2.5. Deep learning features classification

We used a convolutional variation autoencoder (CVAE) to convert the ECG signal into a set of features (also called latent space). The input to this neural network was a QRS signal of dimension (i,300,12), where **i** is the number of 12-lead QRS signals to be processed. CVAE has three system units: encoder, latent space unit, and decoder. An encoder consists of several successive convolutional blocks. Each block consists of a convolution layer, a batch normalization layer, an activation layer and a maximum pooling layer. The encoder block transforms a QRS complex into **k** feature vectors, where **k** is the parameter of the encoder. We used the same encoder unit for the QRS signal of each lead with shared weights. At the output of the encoder, the received features are concatenated into a block of latent space, and the dimension is further reduced by the linear transformation. One-dimensional deconvolution blocks are used to decode the signal. Each block consists of a convolutional transpose layer, a batch normalization layer, and an activation layer.

The loss function for training this neural network is as follows:

$$L = \sum_{n=1}^{12} ||QRS_{true} - QRS_{pred}||^2 + KL[N(\mu,\sigma), N(0,1)]$$

(1)

where QRS_{true} is the input QRS signal in one on the 12 lead, QRS_{pred} - decoded QRS signal, KL - Kullback-Leibler divergence, $N(\mu,\sigma)$ - normal distribution with parameters μ and σ given out from latent space.
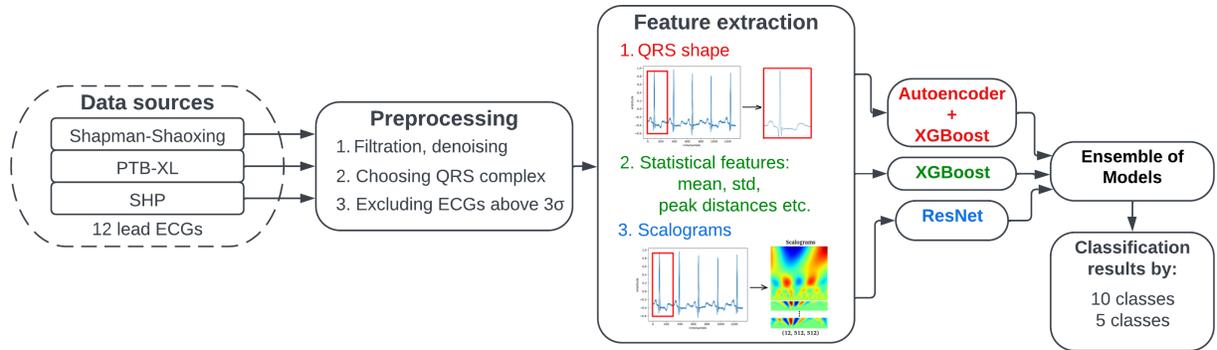
Figure 1. Pipeline for the classification of ECGs.

For ECG classification we used trained CVAE on described datasets and then used features from the latent layer for the ECG classification task. For this, we implement a neural network with several dense layers and a sigmoid layer as an output layer. We used a binary cross-entropy loss function to train this classifier.

2.6. Ensemble of ML models

In our work, we employed ensemble methods, which merge different algorithms and architectures to generate predictions by aggregating their outputs. To combine ML model outputs we utilized the averaging approach, where each model's prediction is given equal weight in the final decision-making process.

3. Results

The results of the ensemble of ML models for 10-class and 5-class classification are shown in Table 1 and Table 2 respectively. For both classifiers, we obtained high values for accuracy, but the considered dataset was imbalanced, so we have focused further description for the F1 score, sensitivity, and specificity. For the **10-class** classifier, we got a value of F1-score higher than 0.7 for Normal ECG, LBBB, LAFB, IRBBB, and RBBB. In addition, we obtained a fairly good ratio ($\geq$ 0.8) of sensitivity and specificity for Normal ECG, LBBB, LAFB, and RBBB. We did not have a high sensitivity (0.72) for IRBBB. This was because a proportion of true IRBBB cases were classified as RBBB and normal ECG. For ILBBB, LBFB, and VPE we obtained zero values for F1 and sensitivity, again due to the imbalance of the dataset and the low presence of these classes in the training and validation sets. For PVCs, we got low values of F1-score and sensitivity. This can be explained by the chaotic appearance of PVCs on ECGs, while we used only one QRS complex by the 12-lead ECG in the classification. As a result, PVCs may not have reached the selected QRS complex. We also got low F1 scores and sensitivity values for the NICD pattern.

ECGs in cases with NICD can vary from patient to patient. This may be due to the presence of fibrotic changes and infarcts that affect the electrical excitation of the heart. Thus, ECGs from patients with fibrosis and block not related to LBBB and RBBB may be attributed to NICD, complicating NICD classification.

For a **5-class** classifier, we achieved a F1-score (≥ 0.7) and a reasonably balanced sensitivity-to-specificity ratio ($\geq$ 0.8) for non-infarcted ECGs, anteroseptal MI, and inferior MI. However, we observed zero values for both the F1-score and sensitivity in the cases of lateral and anterior MI. We attribute this subpar performance for these classes to their limited representation within the training dataset.

4. Discussion and Conclusions

This paper presents the outcomes of employing an ensemble of machine-learning models for ECG-based diagnosis classification. We explored two classifiers, one with 10 classes and another with 5 classes. Our evaluation revealed good accuracy and performance, as measured by the F1-score and the sensitivity-to-specificity ratio, for certain diagnoses. However, we observed suboptimal outcomes for specific diagnoses, primarily stemming from class imbalance issues.

Furthermore, our classifier utilizes a 12-lead ECG as input, with only one QRS complex per lead, which may impact the accuracy of ventricular premature contractions (PVCs) classification. In forthcoming research, we aim to enhance classification outcomes by addressing class imbalances through the generation of synthetic samples.

References

[1] Zhang Z, Dong J, Luo X, Choi KS, Wu X. Heartbeat classification using disease-specific feature selection. Computers in Biology and Medicine 3 2014;46:79–89. ISSN 0010-4825.

[2] Afkhami RG, Azarnia G, Tinati MA. Cardiac arrhythmia classification using statistical and mixture modeling features of ecg signals. Pattern Recognition Letters 1 2016; 70:45–51. ISSN 0167-8655.

Table 1. Results of the ECG classification for 10 classes.

Labels	Training, n	Validation, n	ROC AUC	Accuracy	Specificity	Sensitivity	F1-score
Normal ECG	17899	4484	0.91	0.84	0.85	0.84	0.76
LBBB	569	143	0.96	0.98	0.99	0.87	0.83
ILBBB	41	20	0.83	1.00	1.00	0.00	0.00
PVCs	1549	388	0.79	0.83	0.89	0.51	0.5
LAFB	1362	341	0.98	0.95	0.96	0.86	0.83
LPFB	132	36	0.55	0.98	1.00	0.00	0.00
IRBBB	1747	438	0.90	0.88	0.92	0.72	0.70
RBBB	1223	306	0.93	0.97	0.98	0.86	0.87
NICD	568	142	0.82	0.93	0.97	0.34	0.37
VPE	95	24	0.53	0.99	1.00	0.00	0.00

Table 2. Results of the ECG classification for 5 classes.

Labels	Training, n	Validation, n	ROC AUC	Accuracy	Specificity	Sensitivity	F1-score
Non Infarcted	7738	1963	0.94	0.84	0.85	0.91	0.87
Anteroseptal MI	1860	438	0.94	0.89	0.91	0.80	0.74
Lateral MI	159	41	0.44	0.98	1.00	0.00	0.00
Inferior MI	2455	644	0.93	0.86	0.88	0.82	0.78
Anterior MI	524	147	0.78	0.94	1.00	0.00	0.00

[3] Garcia G, Moreira G, Menotti D, Luz E. Inter-patient ecg heartbeat classification with temporal vcg optimized by pso. Scientific Reports 2017 71 9 2017;7:1–11. ISSN 2045-2322.

[4] Li H, Yuan D, Ma X, Cui D, Cao L. Genetic algorithm for the optimization of features and neural networks in ecg signals classification. Scientific Reports 2017 71 1 2017; 7:1–12. ISSN 2045-2322.

[5] Alotaiby TN, Alrshoud SR, Alshebeili SA, Aljafar LM. Ecg-based subject identification using statistical features and random forest. Journal of Sensors 2019;2019. ISSN 16877268.

[6] Mar T, Zaunseder S, Martínez JP, Llamedo M, Poll R. Optimization of ecg classification by means of feature selection. IEEE Transactions on Biomedical Engineering 8 2011;58:2168–2177. ISSN 00189294.

[7] Jung WH, Lee SG. An arrhythmia classification method in utilizing the weighted knn and the fitness rule. IRBM 6 2017;38:138–148. ISSN 1959-0318.

[8] Krasteva V, Jekova I, Leber R, Schmid R, Abächerli R. Superiority of classification tree versus cluster, fuzzy and discriminant models in a heartbeat classification system. PLOS ONE 10 2015;10:e0140123. ISSN 1932-6203.

[9] Ribeiro AH, Ribeiro MH, Paixão GM, Oliveira DM, Gomes PR, Canazart JA, Ferreira MP, Andersson CR, Macfarlane PW, Wagner M, Schön TB, Ribeiro ALP. Automatic diagnosis of the 12-lead ecg using a deep neural network. Nature Communications 2020 111 4 2020;11:1–9. ISSN 2041-1723.

[10] Lannoy GD, François D, Delbeke J, Verleysen M. Weighted conditional random fields for supervised interpatient heartbeat classification. IEEE Transactions on Biomedical Engineering 1 2012;59:241–247. ISSN 00189294.

[11] Zhang D, Yang S, Yuan X, Zhang P. Interpretable deep learning for automatic diagnosis of 12-lead electrocardiogram. iScience 4 2021;24:102373. ISSN 25890042.

[12] Zheng J, Zhang J, Danioko S, Yao H, Guo H, Rakovski C. A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients. Scientific Data 2020; 7(1):48.

[13] Wagner P, Strodthoff N, Bousseljot RD, Kreiseler D, Lunze FI, Samek W, Schaeffter T. Ptb-xl, a large publicly available electrocardiography dataset. Scientific Data 2020;7(1):154.

[14] Liu H, Chen D, Chen D, Zhang X, Li H, Bian L, Shu M, Wang Y. A large-scale multi-label 12-lead electrocardiogram database with standardized diagnostic statements. Scientific Data 2022;9(1):272.

[15] Arzeno NM, Deng ZD, Poon CS. Analysis of first-derivative based qrs detection algorithms. IEEE Transactions on Biomedical Engineering 2 2008; 55:478–484. ISSN 00189294.

[16] Bostanov V. Bci competition 2003-data sets ib and iib: feature extraction from event-related brain potentials with the continuous wavelet transform and the t-value scalogram. IEEE Transactions on Biomedical Engineering 2004;51(6):1057–1061.

[17] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016; 770–778.

Address for correspondence:

XSpline S.p.A,
2F Josef Ressel, Bolzano, 39100 , Italy
bazhutina@xspline.com